Japan, December 1st 2021

Teaching Statistics and Data Science: an international perspective

Helen MacGillivray
Queensland University of Technology
Australia

Outline

* Data science, data literacy and statistics

* Good practice in teaching statistics and data science and writing about it

* Leadership in practice and teaching of statistics and data science

* What is needed

References

Data science, data literacy and statistics

Data science, data literacy and statistics

- At UN World Data Forum 2019
  Data Science leaders: '*Data Science ….label for work done for years'*
  'teams need diversity but all need statistics'

- Donoho (2015): overwhelming evidence that statistics is at heart of data science
  - Tukey's (1962) '*Data analysis*'; Chambers (1993) '*greater statistics*'; Cleveland's (2001) '*Data science*'
- Donoho's '*greater data science far more than a mere scaling up to big data and big technology*' … ongoing '*more intellectually productive and lasting*' science

- Greater statistics and greater data science for greater data.

- Gould (2021): '*data-scientific thinking has statistical thinking at its core , blended with some computational thinking, and with a dash of mathematics*.'

So little difference…

- Data literacy descriptions are same as long-established statistical literacy
  - Contexts may differ
  - Gould (2017): both types need updating
- Descriptions of data cycle/learning from data are updates of variety of descriptions of data/statistical investigations/investigative cycle
- From statisticians into statistical education
    - PPDAC (Problem, Plan, Data, Analysis, Conclusion) from industrial statistics early 1990's from Tukey (1980)
    - PCPD (Plan, Collect, Process, Discuss) from 1970's
    - Cameron (2009), Kenett & Thyregod (2006)
  - Issues/problem (elicitation, formulation to tackle)
  - Source, consider & prepare data
  - Explore, analyse
  - Bring together, evaluate, revisit
  - Discuss, communicate

Natural diversity and unity of statistics and data science

- Inextricable to all endeavours finding meaning in data, variation and uncertainty.

- The most theoretical to the most applied topic, will have come from, or link to, real contexts and problems in at least one other discipline.
- Rodriguez (2013) described Statistics as the "*most unselfish of sciences*"; and that "*Statistics improves human welfare not by its own ends, but by its contributions in all fields*."

- Statistics has always advanced, and been advanced by, developments in other disciplines, including computer science and maths – both serve statistics.
- 'Data science' is part of natural development of science of data, variation, uncertainty.

- Data science without statistics and statistical thinking is computer science not data science.
- Wobbly internal 'boundaries' within the statistical and data sciences are 'noise'.

It’s time ….

- ‘Big tent’ of Rodriguez (2013) is statistics and data science.
- Statistics and technology have been mutual enablers and enabled
- BUT the ‘big tent’ is now even bigger, and the technology of obtaining, sourcing, evaluating, credentialling, handling, storing, accessing, combining, ‘wrangling’, exploring, presenting …data must be much more present in education and training across levels, curricula and disciplines.

- ‘Data science’ fervour is opportunity
  - to emphasize greater statistics and greater data science
  - to reinvigorate advocacy that statistics learning reflect the practice of statistics
  - to reflect on why such advocacy has not had sufficient penetration or sustained implementation
- The pedagogy of statistical and data enquiry-oriented learning matters more than ever.

- This matters across diversity but most at foundation, introductory, across disciplines

Good practice in teaching statistics and data science and writing about it

- *"Teaching Statistics* is intended for all those who teach statistics and data science to school age students, or to tertiary students at the introductory level, particularly across disciplines. The emphasis is on good practice in teaching statistics, statistical thinking and data science in any context."
  - Published by Wiley for the Teaching Statistics Trust since 1979.
  - Arose from International Statistical Education newsletter for ISI members, an initiative of ISI's Education Committee,1948.
  - Other initiatives led to ICOTS, Roundtables. Committee became IASE in 1992.
  - Individual subscription rate is ONE-THIRD for IASE/NCTM members (e.g. 22 Euros)

- In past 5 years
  - Submissions from 50 different countries
  - Approximately 100 different reviewers worldwide
  - Average just under 1 submission per week; approximately 40% acceptance
  - Editorials started in 2014; in every issue from 2016 – have attracted considerable general attention

Special issue 2021: *Teaching Data Science and Statistics: foundation and introductory*

- Evolved from increasing submissions reflecting:
  - technology in teaching statistics
  - data and contexts considered
  - broadening of statistical issues, explorations, presentations, and discussions
- Invited and contributed – authors from 8 countries
- 19 chapters (225 pages) grouped under following headings

  THE WAY FORWARD

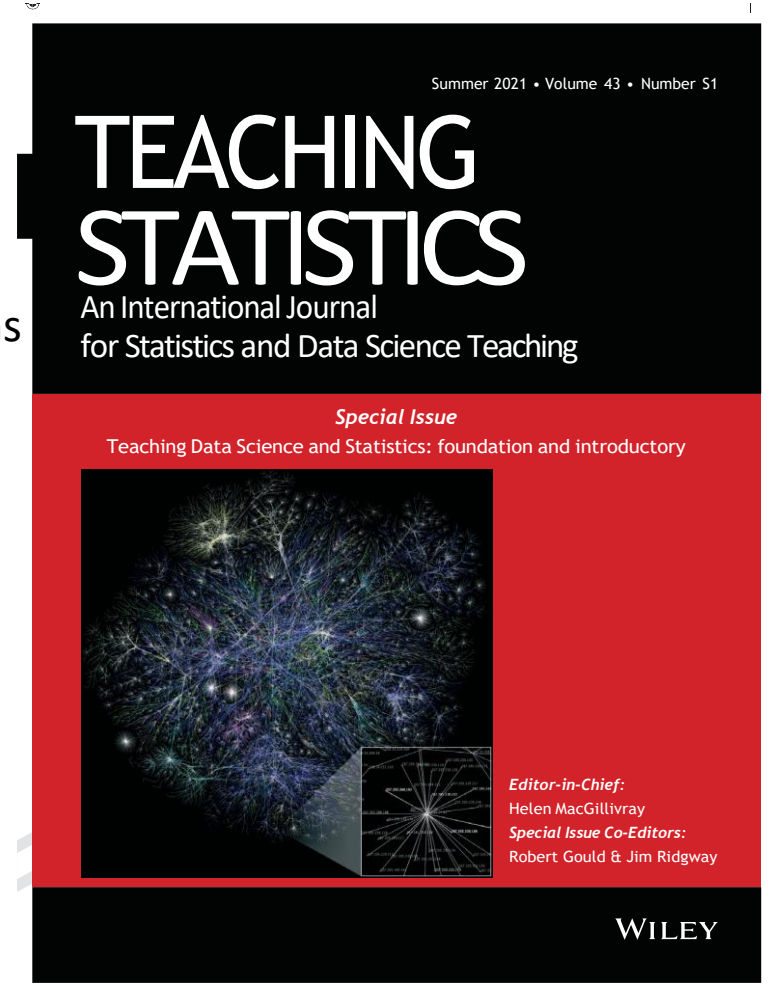
  DATA INVESTIGATIONS IN CONTEMPORARY ISSUES

  EMBEDDING TECHNOLOGY, DATA AND LANGUAGE

  DATA SCIENCE IN CURRICULA: FROM PRIMARY TO WORKPLACE

- 13 items of online supplementary materials

- Editorial "*Statistics and data science must speak together*"



Journal submissions: trends, good facets

- Good facets and trends
  - Less editorial help needed
  - ScholarOne & Wiley support
  - Great reviewers – thorough, constructive, high standards
  - Broadening of teaching contexts – introductory levels at upper ug, pg & workplaces

- More good trends
  - Case studies for teaching
  - Technology for statistics & statistical purposes
  - More complex data & sources, data accessing & wrangling
  - Broader data and contexts considered: multivariable and contextual complexities
  - Student-driven issues and questioning of data sources and quality
  - Broadening of statistical issues and approaches: seeking procedures for data and contexts
  - Visualisation, explorations, discussions

- **Need more submissions in good trends!**

Journal submissions: not so good facets

- Topic, content
  - Trivial, trivialised or negligible statistics value
  - Poor, mistaken (or absent!) statistical understanding, knowledge, procedures
  - Maths for maths' sake
  - Technology for technology's sake
  - Psychology for psychology's sake
  - Assumptions not stated or evaluated
  - No or inadequate visualisation or exploration

- Writing
  - Not international
  - Inadequate (or absent) description of teaching context
  - Too much and 'surface' referencing
  - Not addressing those who teach – hidebound by research styles

Need more on good practice in teaching statistics

- Scholarly writing
  - Teaching context
  - International context
  - Analytical
  - Identify assumptions and evidential sources
    - Statistics education could benefit from looking at statistical investigation methods in areas such as medicine
  - Reference depth not superficial breadth
  - Statistically meaningful – depth and breadth
  - Writing for good practice in teaching greater statistics – and now greater science
    - Uses and drives developments in other disciplines, including maths, computer science, psychology etc, but not driven by them

Need culture change to solve ongoing problems in statistics teaching … statistical practice not sufficiently embedded and ….

- Too much focus on new ways of learning old content & old sequencing
  - Attempts to reproduce math stats textbooks ‘minus’ the maths
- Fixation with restrictions to one and two variables
- Instructor-prescribed question and desired answer
- Multiple procedures and forcing into discipline norms
- Rigid, discipline-embedded approaches, top-down case studies, and too much orientation for research training – in any discipline, including statistics
- Need to understand
  - Types of data: numerical codes cannot turn
    - nominal variables into numerical variables
    - ordinal variables into continuous variables
  - Different visualisations, presentations
  - Aggregation from histograms to maps
  - Assumptions and checking

Most importantly, move to many variables and real empowerment as soon as possible

Opportunity to reinvigorate advocacy and implementation…

- Authentic and embedded learning of statistical/data investigation cycle
- Authentic original contexts and data, with simple embedded in the complex
- Collecting/sourcing data relevant to students' lives.
- Use of technology as in the *practice* of statistics.
- Multivariable contexts and data, moving to multivariable data as quickly as possible.
- Visualisation and exploration.
- Student-centred learning.
- Hands-on learning of data acquisition, handling, wrangling.
- The above applying in all introductory data science learning across disciplines.
- Reclaim and reform the teaching of probability
    - Embed data, language: probabilities assigned/estimated; conditioning language
    - Conditional probability first: many-variable civic data - splitting, confounding, hidden

And opportunity to expand partnerships e.g. official statistics

- United Nations Global Network of Institutions for Statistical Training (UN GIST) set up 2018

- As inaugural Chair of GIST, set up and oversaw task teams on topics ranging from statistical literacy (For whom? By whom?), e-learning, course reviews and meeting the needs of NSO's (National Statistical Offices)
  - UNITAR MOOC demonstrated importance of foundations

- Special issue of SJIAOS on *New Developments in Training in Statistics*
  - Asked to write *Getting the foundations right* (MacGillivray, 2021)

- Again, correspondences between good practice in *teaching* and *doing* statistics

Leadership in practice and teaching of statistics and data science

Leadership in practice of statistics

- Statistical leadership of Gibson (2019): '*use of influence to guide a multidisciplinary research team*'..'*statistician as an individual leader, enterprise leader or policy leader*'

- '*entwined collaboration*' of Cameron (2009) cf '*serial collaboration*' & '*straight consulting*'

- **Collaborative** leadership: leadership in ***professional practice*** **of statistics**

- Gibson (2019): '*active listening, asking questions, networking and communication*'
  - + authentic statistical thinking/understanding = much long-standing advice from statistical consultants

- Barnett (1986) '*we see, tied up together, the role of the statistician as consultant, consultancy as the stimulus for research in statistics, and consultancy as the basis for teaching statistics*'

Collaborative leadership in statistics education/teaching

- Joiner (2005) listed skills need in statistical consulting, including:
- *Be a good teacher—much success in consulting depends on being able to help others understand statistical tools, and their strengths and weaknesses.*

- **Collaborative leadership in practice of statistics same characteristics as in teaching statistics**

- ‘Service’ teaching leadership
- Importance for statistics cannot be emphasized enough

- Extends into joint curricula; school curricula; research methods for advanced or postgraduate students; workplace professional development; advice to colleagues; and to conducting reviews of statistics in a variety of university scenarios.

Collaborative leadership in ‘service’ teaching

- Needs active listening and observation, identification of issues, sourcing and discovering information (data), analysis, synthesis and modelling, communication of findings or product, and evaluation and reflection.
- Listen carefully to other disciplines’ desires and frameworks, find out as much as possible, and transmute this into a realistic model that meets their students’ needs both immediate and long-term.
- This should be explained, justified, evidenced and inspiring to the other discipline(s).

- There is no fixed introductory course in statistics and no fixed content order
- Time, logistics, student background, level, outcomes, practical and financial…. all matter
- Coherence and the statistical story to maintain engagement

- Move quickly and purposefully to the capacity and power of statistics for cohort
- Learning experiences which students own

Collaborative leadership in ‘service’ teaching

- Continue collaboration with other discipline(s), keep them up to date with personal feedback, as well as qualitative (survey) and quantitative (assessment) student data.

- Achieve workload balance through analysis, planning, organisation, collaboration and mentoring of all staff involved.
- It is possible
  - to participate in students hands-on learning experiences across the whole cohort
  - to include open-ended learning experiences and assessment, by cleverly designed preparation and other quiz-type assessment.

- Frankness with students: clear justifiable parameters = freedom within parameters.

- Reviews of curricula in universities, working on school curricula with teachers and educational authorities, and working with business and industry on workplace professional development all need this collaborative leadership and problem-solving process.

Future statisticians and data scientists

- Increasing attention on work-integrated learning, capstone and hands-on learning of statistical consulting skills in advanced undergraduate, postgraduate

But

- Statisticians been advocating this for years and authentic experience of data investigations from start
- Cameron (2009) comments that thorough grounding in this full process is an appropriate foundation for most statisticians wherever they may be employed.

- Future statisticians need embedded in all their courses:
  - experiential and constructivist learning
  - environments facilitating exploring and learning problem-solving
  - statistical thinking in all its manifestations
  - communicating, collaborating, constructing meaning

Communication of statistics to others – how to teach statistical understanding

- Statistical communication embedded throughout courses

- Developmental and mentored program: peer assistance → tutoring

- Comments from volunteers in mentored learning support program:
  - *Fantastic! Love it.*
  - *Brilliant for learning*
  - *1-1 conversations to learn ways of explaining*
  - *Love being able to help with no pressure*

- From graduates in workplace:
- Without these tools the role of the statistician in a workplace where the majority of staff have a minimal statistics background '*would have been near impossible'.* (Gibbons and MacGillivray, 2014)

Collaborative leadership to develop tertiary teaching expertise

- Extensions of these principles to mentoring and supporting early career staff include:
  - swapping roles;
  - sharing resources;
  - team reflections and analysis;
  - adapting resources to other teaching contexts;
  - articulating criteria and standards;
  - designing integrated authentic, effective but efficient assessment;
  - identifying strengths of individuals and strategies;
  - interactive collaborations designed to assist early career staff to develop teaching strengths as well as balance their workloads and further their careers and career satisfaction.

Few words about school level

- Above principles and collaborative leadership model apply

  But
- Contexts very different and contexts of primary, middle senior school differ
- Challenges include:
  - curricula across schools;
    - ‘parcelling up’ learning for school contexts;
    - textbooks;
    - assessment
- *‘In the US, while we have good documents on what should be happening, the reality is far from what those documents suggest - for a variety of reasons, most of the emphasis is on ….. content that is easy to assess’*

- Who should teach statistics and data science content within school curricula?
- Those who have learnt what it is and how to teach it
- Gould (2021) Rubin (2021) & others show how much time this takes

What is needed

What is needed? A cultural shift

- Greater acknowledgement/respect for skills and expertise required for good teaching of statistics and data science in and across all disciplines, especially foundation and introductory. Also for collaborative research by statisticians.
- No matter what the context or discipline, nor how the curricula/teaching is organised, there needs to be input from statistical and teaching statistics expertise
- Leaders must advocate and promulgate principles and practicalities, and respect progress and initiatives in teaching statistics
- Greater respect for responsible work

- University-wide teaching statistics and data science community – no matter how staff are grouped or spread.

- Accreditation of high level expertise in teaching statistics/data science – may identify specialities but must be expertise in teaching practice.

Again, correspondences between good practice in *teaching* and *doing* statistics

- And data science if this is to be genuine ‘learning from data’

- Much work needed and we need to
  - listen and observe – students, practitioners (statisticians, data scientists, teachers), users …. across statistics
  - identify issues
  - source, discover, access, evaluate, wrangle ….. information (‘the data’)
  - analyze
  - synthesize and model
  - communicate findings or product
  - evaluate and reflect

Gather and use evidence from professional statistical/data scientists

- Advocacy of statistical educators has come from statisticians
- Content pedagogy is & must be grounded in such advocacy, including enquiry-oriented learning, and experiential learning of the full data investigation process.
- Cameron (2009) comments that a foundation of active learning starting in first year of the full data investigation process is an appropriate foundation for most statisticians wherever they may be employed, including research.
- Donoho (2015): "*Many of my audience at the Tukey Centennial where these remarks were presented are applied statisticians, and consider their professional career one long series of exercises in …. collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of ... applications*."

Application then theory, not theory then application, applies as much in teaching data science as it should, & should have, in teaching statistics.
Mathematics and computer science serve statistics and data science.

Diversity of leadership

- Inspirational – focussed, breakthrough
  - Connecting with giants
  - Connecting locally and personally

- Professional: in-reach and outreach – deeply & truly responsible and far-sighted
  - Stuart Rice, 1948: ISI; Associations framework; ISI Education Committee
  - Vic Barnett: *Teaching Statistics*, ICOTS, UK Centre for Statistical Education.
  - Community building, mentoring; representation across diversity

- Inclusive in general leadership
  - Bourke and Dillon (2016) '*six signature traits*': *cognizance, curiosity, cultural intelligence, collaboration, commitment, and courage*.
  - I add communication – within and with others – both require respect.

Leadership by all in the ‘big tent’ of statistics and data science

- ‘Commitment’ is to the broad tent of the statistical sciences, to its community, integrity and diverse roles in its contributions to all fields and to human welfare.
- ‘Courage’ is in speaking out, representation and advocacy of the practice and teaching of ‘greater statistics’, and support of other statistical leaders.
- ‘Curiosity’ is active listening, asking questions, and thinking outside the box.
- ‘Cognizance’ is reflecting, analysing, seeing and synthesizing the big picture in statistics and articulating its meaning.
- ‘Cultural intelligence’ is respect, discovery of others’ frameworks and adapting in building authentic, productive and progressive community.

Leadership by all in the ‘big tent’ of statistics and data science

- ‘Collaboration’ is at the heart of all the above and the whole statistical community.
- ‘Communication’ within and across the whole statistical community, with collaborators, with students and with the global community of statistical users and all citizens.

Thank you and here’s to Statistics and Data Science!

References

Barnett, V. (1986) Statistical consultancy – a basis for teaching and research. *The Proceedings IASE/ISI 2nd International Conference on Teaching Statistics*, Vancouver: ISI, Voorburg, The Netherlands. http://iase-web.org/documents/papers/icots2/Barnett.pdf

Bourke, J. and Dillon, B. (2016). The six signature traits of inclusive leadership: thriving in a diverse new world. https://www2.deloitte.com/us/en/insights/topics/talent/six-signature-traits-of-inclusive-leadership.html

Cameron, M. (2009) Training statisticians for a research organisation. *Proceedings of the International Statistical Institute 57th Session*. Durban, South Africa. ISI.

Chambers, J. (1993) Greater or lesser statistics: a choice for future research, *Statistics and Computing* 3: p182-184

Cleveland, W. (2001) *Data Science: an action plan for expanding the technical areas of the field of statistics*. International statistical review. 69: p21-26

Donoho, D. (2017) 50 Years of Data Science, *Journal of Computational and Graphical Statistics* 26 (4): p745-766, DOI: 10.1080/10618600.2017.1384734

Gibson, E. (2019) Leadership in Statistics: increasing our value and visibility. *The American Statistician* 73 (2), 109-116, DOI:10.1080/00031305.2017.1336484

Gould, R. (2017) Data literacy is statistical literacy. *Statistics Education Research Journal* 16: p22-2., http://iase-web.org/Publications.php?p=SERJ

Gould, R. (2021) Towards Data Scientific Thinking. *Teaching Statistics. Special issue on Teaching Data Science and Statistics: foundation and introductory* 43 (SI1): p S11-S22

Kenett, R. and Thyregod, P. (2006) Aspects of statistical consulting not taught by academia. Statistica Neerlandica, 60: p396-411

Joiner, B. (2005) Statistical consulting. In Kotz, S., Balakrishnan, N., Read, C. and Vidakovic, D. (ed) *Encyclopedia of Statistical Sciences* (2nd edn) Wiley.

Rodriguez, R. (2013) Building the Big Tent for Statistics, Journal of the American Statistical Association, 108:501, p1-6, DOI: 10.1080/01621459.2013.771010

Rubin, A. (2021) What to consider when we consider data, *Teaching Statistics. Special issue on Teaching Data Science and Statistics: foundation and introductory* 43 (SI1); p S23-S33

Tukey, J. (1980) We need both exploratory and confirmatory. The American Statistician, 1980; 34: p23-25.

Tukey, J. (1962) The future of data analysis. *The Annals of Mathematical Statistics* 33: p1-67.

For some expansion of discussion see

Forster, M. and MacGillivray, H. (2010). *Student discovery projects in data analysis*. In Reading, C. (ed) The Proceedings IASE/ISI 8th International Conference on Teaching Statistics, Ljubljana: ISI, Voorburg, The Netherlands. http://icots.net/8/cd/pdfs/invited/ICOTS8_4G2_FORSTER.pdf

Gibbons, K. and MacGillivray, H. (2014) *Education for a workplace statistician*. In MacGillivray, H.L., Martin, M. and Phillips, B. (2014) (eds) *Topics from Australian Conferences on Teaching Statistics: OZCOTS 2008-2012,* (pp267-294) Springer Science+Business Media, LLC, New York.

MacGillivray, H. (2021) Leadership across the diversity of statistics. In Amanda L. Golbeck (Ed.), *Leadership in Statistics and Data Science: Planning for Inclusive Excellence.* p387-409. New York City, Springer Nature.

MacGillivray, H. (2021) Statistics and Data Science must speak together. *Teaching Statistics. Special issue on Teaching Data Science and Statistics: foundation and introductory* 43 (SI1): p S5-S10

MacGillivray, H. (2021) Getting the foundations right. In *New Developments in Training in Statistics* Statistical Journal of the IAOS 37 (2021) 735–745. DOI 10.3233/SJI-210859

MacGillivray, H. (1998) *Developing and synthesizing statistical skills for real situations through student projects*, Proc. 5th International Conference on Teaching Statistics, 1149-1155, ISI: Voorburg, Netherlands